

## The Design of Competency-based Performance-Assessment 1

Running head: COMPETENCY-BASED PERFORMANCE-ASSESSMENT

This is a pre-print of the article that was published as

Sluijsmans, D. M. A., Prins, F., Martens, R. (2006). A framework for integrated performance assessment in E-Learning. *Learning Environments Research*, 9(1), 45-66.

Copyright Springer; *Learning Environments Research* is also available  
<http://www.springerlink.com/content/102953/>

© Springer 2006

The Design of Competency-based Performance-Assessment in E-Learning

Dominique Sluijsmans, Frans Prins

Open University of the Netherlands

Rob Martens

Leiden University

Correspondence concerning this paper should be addressed to Dominique Sluijsmans, Open University of the Netherlands, Educational Technology Expertise Centre, P.O. Box 2960 6401 DL Heerlen, The Netherlands, voice: ++31-45-5762178, fax: ++31-45-5762802, e-mail: [dominique.sluijsmans@ou.nl](mailto:dominique.sluijsmans@ou.nl)

### Abstract

Though effort has been invested in designing powerful e-learning environments, relatively little attention has been paid to the design of valid en reliable assessments in these environments. This article focuses on the design of competency-based performance-assessment in e-learning. An approach to designing performance assessments in e-learning contexts is presented, that implies also an alternative view on instructional design, learning and assessment. The requirements for the learning environment are addressed. Examples from a virtual seminar are presented to illustrate the design approach. The article concludes with the identification of possible pitfalls related to the approach and gives directions for future research.

### The Design of Performance-based Peer Assessment in E-Learning

Institutions of higher education are confronted with a demand for competency-based learning (CBL), which is expected to narrow the gap between learning in the educational setting and future workplace performance. In these competency-based learning environments learners are confronted with authentic, open problems and learning materials which have personal meaning for them and are presented in a variety of formats. Teaching methods are applied which arouse interest, activate prior knowledge, clarify meanings, and model appropriate learning strategies and reflective processes. The to-be-acquired knowledge and skills are supposed to be integrated in educational activities, so that learners recognise a resemblance with tasks in the real world (Stoof, Martens, van Merriënboer, & Bastiaens, 2002).

Although many efforts have been made to implement CBL and assessment in face-to-face education, it is still quite a struggle when it comes to the design of CBL in an electronic learning environment. Especially the design of more performance-based assessment is a weak component in e-learning. The emphasis is still much more on traditional testing than on assessment. The Standards for Educational and Psychological Testing (1999) define test as “an evaluative device or procedure in which a sample of an learners’ behaviour in a specified domain is obtained and subsequently evaluated and scored using a standardized process.” Testing is often a process of gathering data and returning results, instead of a process of providing opportunities for learning. An assessment is a more general process of gathering data to evaluate a student. Data from several assessments are used to make a decision about a learner’s performance level.

In contrast to traditional tests, assessments are based on multiple products or processes, for example essays, reflection papers, oral assessments, process analyses, group products, and work samples. The assessment task is described in terms of a certain

performance that is perceived as worthwhile and relevant to the learner, and can therefore be defined as performance assessments (Wiggins, 1989). Performance assessment focuses on the ability to use combinations of acquired skills and knowledge, and therefore fits in well with the theory of powerful learning environments (Linn, Baker, & Dunbar, 1991). Since the goals as well as the methods of instruction are oriented towards integrated and complex curricular objectives, it is necessary for assessment practices to increasingly use various kinds of assessments in which learners have to interpret, analyse and evaluate problems and explain their arguments. These assessments, which are fully integrated in the learning process, provide information about learner progress and support learners in selecting appropriate learning tasks. A shift occurs from a test culture to an assessment culture, which strongly emphasizes integration of instruction, learning and assessment (Biggs, 1996; Birenbaum, 2003). The compatibility between instruction, learning, and assessment, is described within the theory of constructive alignment (Biggs, 1996; see also Birenbaum, 2003). When there is alignment between what teachers want to teach, how they teach, and how they assess, teaching is likely to be more effective than when it is not. To pursue the theory of constructive alignment, it is worthwhile to invest in the design of performance assessments, because performance assessment provides multidimensional feedback for fostering learning (Birenbaum, 2003).

Literature on *testing* in e-learning focuses mainly on tools that are item-based and directed at knowledge-based testing. A major disadvantage of such tests is that they tend to focus on the measurement of low-level retention of isolated facts rather than the application of knowledge to solve ill-structured problems (Baker & Mayer, 1999; Reeves, 2000). Zhang, Khan, Gibbons and Ni (2001), for example, reviewed 10 web-based tools that were all based on the item type testing. One of the well-known applications of these item-based tools in e-learning is computerised adaptive testing (CAT). In CAT, the computer dynamically

evaluates the ability of the student, resulting in a test that is adapted to each individual student. This fine tuning is achieved by statistically adapting the test to the achievement level of each student while avoiding very easy or very difficult questions. Although Zhang et al. (2001) concluded that time and place independency were the main advantages of these tools, they also acknowledge that none of the tools makes use of performance assessments. Web-based tests are yet far away from assessments that support relevant professional performance and learning.

Simulations of hands-on tasks are found useful for more *performance-based assessment* in e-learning (Shavelson, 1991). In the technical area, for example, simulators are developed for certifying pilot competencies (Bennett, 1999), or for training and assessing skills to operate submarine periscopes (Garris, Ahlers & Driskell, 2002). Recent research projects focus on the assessment of problem solving skills (e.g., Mayer, 2002; O'Neil, 2002). Here the computer keeps a record of every move made by the student in solving a task in order to provide a detailed profile of his or her performance for assessment. Video recording is also a realistic option for performance assessment, used in for example teacher education contexts. A video recording can be applied for educational activities as analysis of the observation, peer review or other assignments. To date, ICT is often used as a technology to simulate the context of professional practice (Bastiaens & Martens, 2002). Modern distance learning courses are often set up as realistic 'games' or simulations (e.g., Garris et al., 2002). ICT here is intended to provide the 'virtual' reality as a motivating authentic problem and serves as a provider for competency-based learning. But how to bring the assessment in e-learning in line with these often complex learning environments?

To answer this question, we focus on the design of competency-based performance assessment in e-learning and the implications of integrating performance assessment in e-learning. Assessment that is strongly embedded in instructional practice in general, and in e-

learning in particular, is scarcely out of the egg, so the lack of structured frameworks to guide assessment design is understandable. This article will present design guidelines that support and guide the design of sound performance assessments in e-learning. Throughout the article, examples from a distributed case-based CSCL-course called the *European Virtual Seminar (EVS)*, designed at the Open University of the Netherlands, are presented to illustrate the implementation of competency-based performance-assessment in an e-learning environment as well as possible problems that can be encountered during implementation of performance assessment activities in e-learning.

### Designing Performance Assessments

For the alignment of instruction, learning, and assessment, learning tasks should be directly related to the performance assessment tasks at the end of a study unit. In contrast with the design procedure of most teachers, in which first the learning tasks are designed and then the assessment, Stiggins (1987) states that the design of the assessment should be the first step in educational design. For this, he formulated four general guidelines to design performance assessments (see Table 1).

\*\*\*\*\*INSERT TABLE 1 ABOUT HERE\*\*\*\*\*

First, the *purpose* of the performance assessment has to be defined. Several important questions are in order (Herman, Aschbacher, & Winters, 1992): What important cognitive skills or attributes do students have to develop? What social and affective skills or attributes do students have to develop? What metacognitive skills do students have to develop? What types of problems do they have to be able to solve? What concepts and principles do students have to be able to apply? This first step results in a skill decomposition in which the relevant skills are hierarchically ordered (van Merriënboer, 1997). For example, in the case-based CSCL-course *European Virtual Seminar (EVS)*, the course designers and teachers described the purpose of the performance assessment as follows: “After having participated in the EVS

course, students should be able to (1) describe the concept of sustainable development, (2) give an overview of the relation between EU-enlargement policy-related issue and sustainable development, (3) make a link between regional, national, and European issues concerning sustainable development, (4) work with students with different disciplinary and cultural backgrounds and exchange views on sustainable development in Europe, (5) participate effectively in a computer conference, and (6) use the computer conference for collaborative learning purposes.” As can be seen, these purposes are still rather general. Moreover, this e-learning course illustrates the frequently stated wish of teachers of CSCL courses that students have to acquire and use collaborative learning skills in their course. However, many times students do not get the opportunity to practice and learn these skills.

When the purpose of the assessment is defined, decisions are made concerning the *performance assessment task*. The performance assessment task may be a product, behaviour or extended written response to a question that requires the student to apply critical thinking skills. Some examples of performance assessment tasks include written compositions, speeches, and research projects. It is important that the performance assessment task can be performed in an electronic learning environment, if you want your students to take the task from their computer. In our example, the EVS course, groups of students had to write a report on sustainable development and enlargement of the European Union. This report had to contain useful advice for the European Committee concerning policy of water management, agriculture, or energy.

After the assessment task is determined, the elements of the task that determine the quality of the student's performance are defined. Sometimes, these can be found in so-called job profiles. Although these resources may prove to be very useful, they often include *lists of criteria* that may include too many skills or concepts or may not fit exactly. Most of the time, teachers develop their own criteria. A teacher has to analyse skills or products to identify

performance criteria upon which to judge achievement. This is not easy. It is useful to use expert products or good examples to define the appropriate criteria. Communicating information about performance criteria provides a basis for the improvement of that performance. Quellmalz (1991) offers a set of specific guidelines for the development of quality performance criteria. Criteria should be significant, specifying important performance components, and represent standards that would apply naturally to determine the quality of performance when it typically occurs. The criteria must be clear and understandable for all persons involved. In e-learning environments, the teacher can determine these criteria in negotiation with students in for example discussion groups.

The final step is the creation of *performance scoring rubrics* that are used for formative and summative purposes. Contrary to more traditional forms of testing, performance assessments in which the students often are confronted with ill-defined problems, do not provide clear-cut right or wrong answers. The performance is evaluated in a way that allows for informative scoring on multiple criteria. In a performance scoring rubric, the different levels of proficiency for each criterion should be defined. Using the information of the assessment form, feedback is given on a student's performance either in the form of a narrative report or a grade. A criterion-referenced qualitative approach is desirable, whereby the assessment will be carried out against the previously specified performance criteria. An analytic or holistic judgement then is given on the basis of the standard the student has achieved on each of the criteria. Analytic or holistic rubrics that specify clear benchmarks of performance at various levels of proficiency also serve the purpose of guiding students as they perform the task (Birenbaum, 2003). Nystrand et al. (1993) and Pitts et al. (2001) investigated whether it is preferable to have a holistic approach in performance assessment. When competencies are assessed through a task that requires the learners to integrate them, 'holistic' or 'integrated' assessment is required. This form of assessment relates to the whole unit or



grouping of units, and requires observation of performance, questioning, and in some cases, review of documentation or other forms of evidence.

The performance scoring rubric used in the EVS course is presented in Table 2 (adapted from Prins, Sluijsmans, & Kirschner, 2005). For the summative assessment of the group report and the collaboration process, one of the EVS teachers designed two performance scoring rubrics together with the first two authors of this article. These rubrics were evaluated by the five other teachers and the co-ordinator of the EVS course. The scoring rubric for the product counted for 70% of the end mark (10 criteria, see Table 2) whereas the scoring rubric for the group process counted for 30%. Criteria for the group process, not shown in Table 2, were 1) planning research, 2) planning individual tasks, 3) co-operation within the group, 4) co-operation via the internet, 5) participation, and 6) incorporate comments.

\*\*\*\*\*INSERT TABLE 2 ABOUT HERE\*\*\*\*\*

#### Assuring Quality in Performance Assessment

In CBL, it is important that a number of performance assessments are organised to gather reliable and valid information about a learner's competency-development. The standard approaches to reliability and validity are derived from a psychometric approach, which is based upon the notion of reaching an 'ideal' grade (Johnston, 1994). In competency-oriented learning contexts, Dierick, Watering and Muijtjens (2002) indicate a shift from psychometric to edumetric criteria for the quality of assessment scores. There is more attention for criteria like accuracy of the scores, the cognitive complexity of the performance task, the authenticity of the performance task, the transparency of the assessment procedure, and the fairness in assessment. Each performance assessment however has a weak link in the quality chain that links the performance of the learner and the conclusion about the competency in a particular context (Crooks et al., 1996). To tackle this problem, three

important criteria are specifically important to cover in the design of performance assessments: accuracy, generalizability, and extrapolation.

A performance assessment is *accurate* when the score comes close to the true score of the learner. The true score is a theoretical concept defined as the mean score of an infinite number of measurements by means of equivalent assessments, assuming that there are no changes in the person or any other effects. The assessment of competencies also implies more than one observed performance. The learner has to perform a similar type of tasks in a variety of situations under the same conditions. Studies generally conclude that the *generalizability* of performances in similar tasks is limited (Linn, Baker, & Dunbar, 1991). The main reason for this finding is that the assessments in current learning environments are a poor reflection of all possible tasks that in fact could be presented to the learner (probably due to lack of time and money). It is therefore recommended to define a variety of performance assessments that represent a certain level of authenticity (Gulikers, Bastiaens, & Martens, 2005). *Extrapolation* implies that the attained score reflects the performance level that the learner would achieve in a real working situation. Sometimes this is no problem, because the assessment task does not deviate from the task in the real situation. But often it is. For example when the performance task is too expensive (launch of a Patriot rocket), too dangerous (defuse a bomb), or when the situation is unlikely to occur in real life (the arrest of an armed criminal in a shopping centre). In most assessments the level of realism (i.e., “fidelity”) is reduced. The more the fidelity is reduced, the more difficult it is to prove that the attained score is a realistic reflection of the authentic performance in the working field.

The three quality criteria put heavy demands on the design and organization of performance assessments, but they are also problematic in the sense that optimising one criterion leads to an impairment of another criterion. Therefore, it is important to choose for a design approach for learning and assessment that warrants for all the quality aspects

(Straetmans & Sanders, 2001). A model that provides guidelines to design competency-based education, in which instruction, learning, and assessment are fully aligned, is the Four Component Instructional Design-model, (4C/ID-model), originally developed by Van Merriënboer and others (Van Merriënboer, Jelsma, & Paas, 1992).

#### Instructional Design for Competency-based Performance Assessment

In the 4C/ID-model competencies are defined as complex skills, consisting of integrated sets of constituent skills with their underlying knowledge structures and attitudes (Van Merriënboer, 1997). Examples of complex skills are giving a training (consultant), designing a house (architect), or supervising a public domain (police officer). The basic components of the model are presented in Figure 1. To illustrate the model and the organization of performance assessment and portfolio assessment, we pursue the example of the EVS-course.

\*\*\*\*INSERT FIGURE 1 ABOUT HERE\*\*\*\*

The *tasks* (first component) are the backbone of every educational program aimed at the acquisition of competencies (see Figure 1, which represents the tasks as circles). The tasks are typically performed in a real or simulated task environment and provide ‘whole-task practice’: ideally, they confront the learners with all constituent skills that make up the whole competency.

Learners will typically start their study on relatively simple tasks and progress towards more complex tasks. Complexity is affected by the amount of skills involved, the amount of interactions between skills, and the amount of knowledge necessary to perform these skills. Task classes are used to define simple-to-complex categories of tasks and to steer the process of selection and development of suitable tasks (see the dotted lines around the circles in Figure 1). Tasks within a particular task class are equivalent in the sense that the tasks can be performed on the basis of the same body of knowledge. The basic idea is to use a whole-task

approach where the first task class refers to the simplest version of whole tasks that professionals encounter in the real world. For increasingly more complex task classes the assumptions that simplify task performance are relaxed. The final task class represents all tasks, including the most complex ones that professionals encounter in the real world.

Once the task classes are defined, the tasks can be selected and/or developed for each class. The cases that are selected for each task class form the basis for the to-be-developed tasks. For each task class, enough cases are needed to ensure that learners receive enough practice to reach mastery. It should be noted that the cases or tasks within the same task class are not further ordered from simple to complex; they are considered to be equivalent in terms of difficulty. A high variability of the tasks within the same task class is of utmost importance to facilitate the development of generalized cognitive schemata and reach transfer of learning (e.g., Gick & Holyoak, 1983; Paas & van Merriënboer, 1994).

While there is no increasing difficulty for the tasks within one task class, they do differ with regard to the amount of support provided to learners. Much support is given for tasks early in each task class, which therefore are labelled as learning tasks, and this support diminishes until no support is given for the final learning task in a task class (see the filling of the circles in Figure 1). The last, unguided and unsupported tasks in a task class (i.e., the empty circles) are therefore suitable performance assessment tasks (see also Figure 1). This task is designed according to guidelines of Stiggins (1987) as outlined in the previous. The assessment task focuses on the ability to use combinations of acquired skills, knowledge, and attitudes and therefore fits in well with the theory of competency-based learning environments (Linn, Baker, & Dunbar, 1991).

Obviously, learners need information in order to work fruitfully on learning tasks and to genuinely learn from those tasks. This *supportive information* (second component) provides the bridge between what learners already know and what they need to know to work

on the learning tasks. It is the information that teachers typically call ‘the theory’ and which is often presented in study books and lectures. Because the same body of general knowledge underlies all learning tasks in the same task class, and because it is not known beforehand which knowledge is precisely needed to successfully perform a particular learning task, supportive information is not coupled to individual learning tasks but to task classes (see the ‘supportive information’ in Figure 1).

Whereas supportive information pertains to the non-recurrent aspects of a complex skill, *procedural information* (third component) pertains to the recurrent aspects, that is, constituent skills of a competency that should be performed after the training in a highly similar way over different problem situations. Procedural information provides learners with the step-by-step knowledge they need to know in order to perform the recurrent skills. They can be in the form of, for example, directions teachers or tutors typically give to their learners during practice, acting as an ‘assistant looking over your shoulder’ (ALOYS), information displays, demo’s or feedback. Because procedural information is identical for many tasks, which all require the same recurrent constituent skills, it is typically provided during the first learning task for which the skill is relevant (see ‘procedural information’ in Figure 1).

Finally, if a very high level of automaticity of particular recurrent aspects is required, the learning tasks may provide insufficient repetition to provide the necessary amount of practice to reach this. Only then, it is necessary to include additional part-task practice (fourth component) for those selected recurrent aspects in the training program (see ‘part-task practice’ in Figure 1).

When learners work on an assessment task in a particular task class, they have to show their progress on both the recurrent aspects of performance, which are routines that are consistent from problem to problem situation, and the non-recurrent aspects of performance, which involve problem solving or reasoning and vary over situations.

\*\*\*\*\* INSERT FIGURE 2 ABOUT HERE \*\*\*\*\*

Figure 2 depicts how performance assessment can be intertwined with the four components. In general, different assessment methods are appropriate for each of the components (see van Merriënboer, 1997). The performance scoring rubric can be a valuable tool to provide formative feedback to students. For summative assessment, the performance scoring rubric helps the teacher to reach a balanced grading and final decision. It is up to the teacher or designer to decide if each assessment task (see Figure 1) is used for summative assessment or, for example, only the task at the end of the most complex task class.

From a theoretical viewpoint, assessment of whole-task performance is the only form of assessment that is unconditionally required in integrated e-learning or any other educational setting for complex learning. The 4C/ID model states that students cannot satisfactorily perform such whole assessment tasks if they do not possess the mental models and cognitive strategies (i.e., theoretical knowledge) that help them to perform the non-recurrent aspects of the task and the procedures or rules that govern the performance of the recurrent aspects of the task. Nonetheless, additional assessment of theoretical knowledge may be applied for a number of reasons. First of all, it may help to diagnose students' conceptual problems or misconceptions and yield the necessary information to give them formative feedback for overcoming these problems. And furthermore, it may be used to corroborate the findings from the assessment of whole-task performance, making the whole assessment more reliable.

Like the assessment of theoretical knowledge, the assessment of part-task performance on single recurrent skills may also be considered as an additional element in the whole assessment system. Preferably, the same tools that are used for part-task practice are also used for the assessment of the recurrent skill under consideration. Most drill-and-practice computer programs (e.g., for using grammatical rules in second language learning; applying safety

procedures in industry, or operating particular software packages in business) indeed assess students on their accuracy and speed, and use this information to diagnose errors; to indicate to students that there is an error, and to provide hints that may help students to get back on the right track. Thus, for collecting evidence concerning theoretical knowledge or recurrent skills, traditional tests may be helpful.

When we return to our example, the EVS course we noticed that educational practice has limitations. According to the 4C/ID-model, the course should contain separate performance tasks and several learning tasks with varying guidance and support. However, writing a report in collaboration is often laborious and complex for students, thus, in the EVS course students were requested to write just one report. This suggests that they had no opportunities to practice this competency. The course designers, however, partly removed this draw-back by asking the student groups to first hand in a draft of the report. The student groups received feedback on this draft and, at the end of the course, they could hand in a revised, final draft. The final draft was used for the summative performance assessment. On the other hand, we could also look at it from a curriculum perspective. One can expect that at the beginning of a curriculum, students are able to learn how to write reports in learning tasks with some guidance and support. That allows for a course with a single performance task with formative assessment further on in the curriculum. As for collaborative skills, however, we do not know many courses at the beginning of curricula that allow for deliberate practice on collaborative skills.

Concluding, we argue that one should always focus on performance assessment of whole tasks. The definition of those assessment tasks early in the design process may also be helpful to the development of appropriate learning tasks that guide students towards the assessment task(s) at the end of a task class. Furthermore, one may consider including

additional assessments for theoretical knowledge and for recurrent skills that have been separately practised.

### The Role of Students in Assessment

Especially in e-learning, learners can play a valuable part in performance assessment by means of peer assessment. Peer assessment implies that students evaluate the performances of peers and provide constructive feedback (Sluijsmans, 2002). At least three arguments plead for implementation of peer assessment in e-learning. First, integrating peer assessment supports students with their development into competent professionals who continuously reflect on their behaviour and their learning. There seem to be several ways in which students can be involved in assessment: students can have a role in the choice of performance assessment tasks and in discussing assessment criteria (Mehrens, Popham & Ryan, 1998). Second, it is substantiated that peer assessment promotes integration of assessment and instruction, seeing the student as an active person who shares responsibility, reflects, collaborates and conducts a continuous dialogue with the teacher. Third, peer assessment can decrease the workload of teachers.

In the EVS course, also peer assessment was integrated in the tasks. When the student groups had to hand in the first draft of their report, groups exchanged their reports with another group and evaluated the product of the fellow group according to the performance scoring rubric shown in Table 2. This way, the student groups received feedback from their peers instead of from their teachers. To make it a real interaction, student groups were asked to reply to their fellow group and to indicate how they processed the feedback during revision of the draft.

### Discussion

This article focuses on the integration of instruction and performance assessments in e-learning. An approach for designing competency-based instruction and performance



assessments with a specific focus on in e-learning contexts is presented, which implies an adjusted view on instructional design, learning and assessment. It is argued that one should always focus on performance assessment of whole tasks. The definition of those assessment tasks early in the design process is helpful for the development of appropriate learning tasks that guide students towards the assessment task(s) at the end of a study unit.

However, some possible pitfalls are conceivable, which may occur when working with this approach. This discussion shall briefly address these issues. Specific conditions for a successful implementation of assessment in e-learning, have to be met. The core of this applies to many e-learning situations: students often don't do the things designers or teachers expect them to (Martens, 1998; Lockwood, 1995, Mudrack, & Farrell, 1995; Strijbos, Martens & Jochems, 2004; Jochems, van Merriënboer & Koper, 2004), and it is of great concern. No matter how high the quality of an assessment procedure is, it is no assurance that students learn in the intended way. There is a distinction between 'what is meant to happen', that is, the curriculum stated officially by the educational system or institution, and what teachers and learners actually do and experience 'on the ground', a kind of de facto curriculum. Snyder (1973) labels this the 'hidden curriculum'. In a lab researchers can ask students to read texts, but in 'real life' students have their own hidden curriculum, "adopting ploys and strategies to survive in the system" (Lockwood, 1995, p. 197). Solution to this problem may be the improvement of students' intrinsic motivation. Ryan & Deci (2000) put forward a model that indicates that certain aspects of the social environment and task environment influence student motivation. Relatedness or trust in peers for example and stimulating students to work on assessment tasks with authenticity that are strongly related to the learning tasks can influence intrinsic motivation, according to Ryan & Deci (2000) and Furrer & Skinner (2003).

Even when teachers and educational developers manage to solve motivational problems, there are more possible problems that need to be solved. A risk factor of integrated

performance assessment in an e-learning setting is that the design of these assessments put heavy demands on teachers and developers (e.g., Beijaard, Verloop, Wubbels and Feiman-Nemser, 2000; Sluijsmans, Moerkerke, Dochy & van Merriënboer, 2001). Introducing e-learning and performance assessment is difficult, requires new teacher roles, requires them to collaborate with many stakeholders and can be time consuming. Moreover, students need to be convinced of the usefulness of competency-learning contexts. If students are not convinced of the usefulness of performance assessment and are not sufficiently intrinsically motivated, it is unlikely that the performance assessment will become a success. The perception of the learning environment might play a mediating role in the interplay between students' intended study strategies, their perceptions of the assessment demands and their actual learning strategies (Nijhuis, Segers, & Gijsselaers, 2005).

Still, we believe that, in spite of these possible risks and problems, the benefits of competency-based performance assessment are too large to dissuade from the implementation of performance assessment in e-learning. In general, the implementation of performance assessments in e-learning holds a number of important advantages (Surgue, 2002). The advantages are related to the integration of assessment and instruction, the possibilities for adequate feedback, the involvement of students, and the authenticity of performance assessments. When looking at the implementation in e-learning, Baker and Mayer (1999) state that computers can have a three-fold value in web-based performance assessment. First, computers have the ability to capture process differences. It is possible to trace back indicators that provide information about which thinking processes contributed to a particular performance. Second, computers can make complex processes visible. And third, online scoring and feedback can be provided based on fixed moments or based on a student model. A student model is based on the logging of actions that students conduct during their learning.

The advantages of e-learning that are often mentioned, such as ease of distribution, timeliness, immediate feedback, variety of delivery modes, tracking, long-term cost savings, convenience, are mostly true for item-based tests, but less applicable for competency-based performance assessments, where the benefits are predominantly based on educational grounds. Item-based tests can be valuable for assessment of part-task practice, but are not useful for whole task assessment.

In line with Birenbaum (2003), we conclude that much more research is required to better understand the nature of competency-based performance-based in e-learning and the impact on learning. For instance the negotiating process whereby assessment criteria are set, the process by which students come to internalise the standards of good performance and the impact of motivational processes in general. Also, the role of teachers, and the complex process of curriculum redesign need to be addressed (Sluijsmans, 2002). It is our belief that performance assessment is a crucial factor in educational innovation, which to date too often fails. When students are really motivated to perform, study, learn and collaborate in a new way, and if learning goals and learning processes are much more in line, educational problems such as amotivation, early drop-out, and test behaviour may be decreased.

#### References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Baker, E. L., & Mayer, R. E. (1999). Computer-based assessment of problem solving. *Computers in Human Behavior*, 15, 269-282.

- Bastiaens, Th., & Martens, R. (2000). Conditions for web-based learning with real events. In B. Abbey (Ed.), *Instructional and cognitive impacts of web-based education* (pp. 1-32). Hershey/London: Idea Group Publishing.
- Bennet, R. E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice*, 18, 5-12.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364.
- Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 13-36). Dordrecht: Kluwer Academic Publishers.
- Beijaard, D., Verloop, N., Wubbels, Th. & Feiman-Nemser, S. (2000). The professional development of teachers. In R.J. Simons, J. van der Linden, & T. Duffy (Ed.), *New Learning* (pp. 261-274). Dordrecht, The Netherlands: Kluwer Academic Press.
- Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht The Netherlands: Kluwer Academic Publishers.
- Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education*, 3, 3, 265-285.
- Dierick, S., Watering, G. van de, & Muijtjens, A. (2002). De actuele kwaliteit van assessment: ontwikkelingen in de edumetrie [Current quality in assessment: Developments in edumetrics]. In: F. Dochy, L. Heylen, & H. van de Mosselaer (Red.), *Assessment in onderwijs: nieuwe toetsvormen en examinering in het studentgericht onderwijs en competentiegericht onderwijs* (pp. 91-122). [Assessment in education:

new modes of assessment in student-centred and competency-based education].

Utrecht: Lemma.

- Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in Children's academic engagement and performance. *Journal of educational psychology*, 95, 148-162.
- Garris, G., Ahlers, R., & Driskell, J. (2002). Games, motivation, and learning. *Simulation & Gaming*, 33, 441-467.
- Gick, M., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.
- Gulikers, J., Bastiaens, Th., & Martens, R. (2005). The surplus value of an authentic learning environment. *Computers in Human Behavior*, 21, 509-521.
- Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Jochems, W., van Merriënboer, J., & Koper, R. (2004). *Integrated e-learning: Implications for pedagogy, technology & organization*. London: RoutledgeFalmer.
- Johnston, B. (2004). Summative assessment of portfolios: an examination of different approaches to agreement over outcomes. *Studies in Higher Education*, 29, 395-412.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20, 8, 15-21.
- Lockwood, F. (1995). Students' perception of, and response to, formative and summative assessment material. In F. Lockwood (Ed.), *Open and distance learning today*, pp. 197-207. London: Routledge,.
- Martens, R.L. (1998). *The use and effects of embedded support devices in independent learning*. Ph.D. Thesis. Utrecht, the Netherlands: Lemma BV.

- Mayer, R.E. (2002). A taxonomy for computer-based assessment of problem solving. *Computers in Human Behavior*, 18, 623-632.
- Mehrens, W.A., Popham, W.J., & Ryan, J.M. (1998). How to prepare students for performance assessments. *Educational Measurement: Issues and Practice*, 17, 18-22.
- Mudrack, P. E., & Farrell, G. M. (1995). An examination of functional role behaviour and its consequences for individuals in group settings. *Small Group Research*, 26, 542-571.
- Nijhuis, J., Segers, M., & Gijselaers, W. (2005). Influence of redesigning a learning environment on student perceptions and learning strategies. *Learning Environments Research, an international journal*, 8, 1.
- Nystrand, M., Cohen, A. S., & Dowling, N. M. (1993). Addressing reliability problems in the portfolio assessment of college writing. *Educational Assessment*, 1, 53-70.
- O'Neil, H. F. (2002). Perspectives on computer-based performance assessment of problem solving. *Computers in Human Behavior*, 15, 255-268.
- Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive load approach. *Journal of Educational Psychology*, 86, 122-133.
- Pitts, J., Coles, C., & Thomas, P. (2001). Enhancing reliability in portfolio assessment: 'shaping' the portfolio. *Medical Teacher*, 23, 351-365.
- Prins, F. J., Sluijsmans, D. M. A., & Kirschner, P. A. (2005). Peer assessment in a computer supported collaborative learning environment: a case study. *Assessment and Evaluation in Higher Education*, 30, 417-444.
- Quellmalz, E. (1991). Developing criteria for performance assessments: The missing link. *Applied Measurement in Education*, 4, 319-332.
- Reeves, T.C. (2000). Alternative assessment approaches for online learning environments in higher education. *Journal of Educational Computing Research*, 23, 101-111.

- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well being. *American Psychologist*, 55, 68-78.
- Shavelson, R. J. (1991). Performance assessment in science. *Applied Measurement in Education*, 4, 347-362.
- Slavin, R. E. (1990). *Cooperative learning: Theory, research and practice*. Hillsdale, NJ: Prentice Hall.
- Sluijsmans, D. M. A. (2002). *Student involvement in assessment. The training of peer assessment skills*. Unpublished doctoral dissertation Heerlen, the Netherlands: Open University of the Netherlands.
- Sluijsmans, D., Moerkerke, G., Dochy, F., & van Merriënboer, J. (2001). Peer assessment in problem based learning. *Studies in Educational Evaluation*, 27, 2, 153-173.
- Snyder, B. (1973). *The hidden curriculum*. Cambridge, MA: The MIT Press.
- Strijbos J. W., Martens, R. L. & Jochems, W. M. G. (2004). Designing for interaction: six steps to designing computer supported group based learning. *Computers & Education*, 42, 403-424.
- Strijbos J.W., Martens, R., Jochems, W., & Broers, N. (2004). The effect of functional roles on group efficiency. *Small Group Research*. 35, 195-229.
- Stiggins, R. (1987). Design and development of performance assessment. *Educational Measurement: Issues and Practice*, 6, 33-42.
- Stoof, A., Martens, R., van Merriënboer, J., & Bastiaens, Th. (2002). The boundary approach of competence: a constructivist aid for understanding and using the concept of competence. *Human Resource Development Review*, 1, 345-365.
- Surgue, B. (2002). Performance-based instructional design for e-learning. *Performance Improvement*, 41, 45-50.

Van Merriënboer, J.J.G. (1997). *Training complex cognitive skills*. Englewood Cliffs, NJ:

Educational Technology Publications.

Wiggins, G. (1989) A true test: Toward a more authentic and equitable assessment. *Phi Delta*

*Kappan*, 70, 703-713.

Zhang, J., Khan, B.H., Gibbons, A.S., & Ni, Y. (2001). Review of web-based assessment

tools. In B.H. Khan (Ed.), *Web-based training*, (pp. 287-295). Englewood Cliffs, NJ:

Educational Technology Publications.



Table 1

## A Step-by-Step Approach for Designing Performance Assessments

Step	What to do?
Define the purpose of the assessment	List the skills and knowledge you wish to have students learn as a result of completing a task.
Define performance assessment tasks	Design a performance task which requires the students to demonstrate these skills and knowledge.
Define performance criteria	Develop explicit performance criteria which measure the extent to which students have mastered the skills and knowledge.
Create performance scoring rubrics	Use one scoring system or performance rubric for each performance task. The performance criteria consist of a set of score points which define in explicit terms the range of student performance.

Table 2

*Part of a Performance Scoring Rubric for Writing a Report on Sustainable Development (SD)*

Criteria	Above Standard	At Standard	Below Standard	Points Earned
	<b>10 - 9</b>	<b>8 - 6</b>	<b>5 - 0</b>	
Sustainable development is made operational	Students give a definition used in their report and give practical tools to measure their solutions on this points	Students give a definition used in their report but do not give practical tools to measure their solutions on this points or vice versa	Students do not give a definition used in their report and do not give practical tools to measure their solutions on these points	<b>/10</b>
	<b>5</b>	<b>4 - 3</b>	<b>2 - 0</b>	
The different aspect of SD are used in coherence	The ecological, social and economic aspect of Sustainable Development are used in coherence and balance. Arguments are given for priority	Not all aspects of sustainable development are used, but the ones that are used, are in balance and coherent	The different aspects of sustainable development are not used coherently nor balanced	<b>/5</b>
	<b>10- 9</b>	<b>8 - 6</b>	<b>5 - 0</b>	
Relation problem definition – analysis – solution	Scientific quality of report and logical forthcoming of conclusions and recommendations. Questions asked in the beginning are answered	Scientific quality of report and logical forthcoming of conclusions and recommendations. Not all questions asked in the beginning are answered	Low scientific quality of report and conclusions and recommendations do not come from the chapters in the report. Not all questions asked in the beginning are answered	<b>/10</b>
	<b>10 - 9</b>	<b>8 - 6</b>	<b>5 - 0</b>	
Summary (separately!)	A 2 - 3 page summary is added, with: Background research, recommendations, target group, possible implementation route. The summary is sharp, and provocative	Summary is lacking one of the four points mentioned or leaves room for interpretation	Summary is lacking for two or more of the four points mentioned and leaves room for interpretation. Or no summary is added at all.	<b>/10</b>

Figure Caption

*Figure 1.* The four components in the 4C/ID-model

*Figure 2.* A framework for designing performance assessment in integrated e-learning



